

SPECIFICATION

NUMERICAL INFORMATION RETRIEVING DEVICE

FIELD OF THE INVENTION

本発明は、自然言語で数値情報を検索する数量表現検索装置に関する。

BACKGROUND OF THE INVENTION

自然言語で数値情報を検索できるようにするには、検索装置が数量的な表現に対応できなければならない。

従来の数値情報検索装置では、例えば特開平7-200583号公報に記載されえた技術のように、文脈情報と時刻／数量表現の基準と範囲に関する情報に着目することによって、相対的な表現を含んだ数量／時間表現を入力文として扱うことができる。このような従来の数値情報検索装置では、例えば、「1990年」に対して「翌年」とは「1991年」とであると補完される。同様に、「第1位から第3位まで」に対して「次とは「第4位」とであると補完される。

しかしながら、上記従来の数値情報検索装置では、単位のような表記や基準が多様な数値的な表記や、基準が多様な数値的な表現には対応することができない。

例えば、「100円増加」は、「1000円」に対する「100円増加」であれば単純に「+100」でよいが、「1メートル」に対して「1センチメートル増加」や「1mm増加」であれば、まず数値的な変換をしなければ補完することができない。

本発明は、このような従来の課題を解決することで、利用者が表記や単位系を意識しなくても数値情報を検索することができる数値情報検索装置を提供することを目的とするものである。

SUMMARY OF THE INVENTION

上記課題を解決するために、本発明の数値情報検索装置は、自然言語で数値情報を検索する数値情報検索装置において、検索対象の文書や検索したい数量表現を入力する入力手段、属性を表す属性名と属性を付与するための条件とが格納される属性情報辞書、上記入力された文書や数値表現に上記属性情報辞書を参照して属性を付与する属性付与手段、属性を表す属性名と変換前文字列と変換後文字列とが格納された変換情報辞書、及び上記属性付与後の文書や数値表現に対して、上記変換辞書情報を参照して、属性付与部分の数量を変換前文字列から変換後文字列に変換することによって、属性付与部分の数値情報を変換する数量変換手段と、を備えたことを特徴とするものである。

BRIEF DESCRIPTION OF THE DRAWINGS

図1は、本発明の第1の実施例の、数値情報検索装置のブロック構成図である。

図2は、図1の数値情報検索装置の動作を説明するフローチャートである。

図3は、図2のステップ202の、投入処理の動作を説明するフローチャートである。

図4は、図2のステップ203の、検索処理の動作を説明するフローチャー

トである。

図 5 は、図 1 の属性情報辞書の構成例を示した図である。

図 6 は、図 1 の変換情報辞書の構成例を示した図である。

図 7 は、本発明の第 2 の実施例の、数値情報検索装置のブロック構成図である。

図 8 は、図 7 の数値情報検索装置における、検索処理の動作を説明するフローチャートである。

図 9 は、図 7 の抽出情報辞書の構成を示す図である。

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS OF THE INVENTION

第 1 の実施例

図 1 は本発明の第 1 の実施例の、数値情報検索装置のブロック構成図である。この第 1 の実施例の数値情報検索装置は、入力手段 1，属性付与手段 2，属性情報辞書 3，数量変換手段 4，変換情報辞書 5，文書格納検索手段 6，文書データベース 7，抽出手段 8，及び出力手段 9 を備える。

入力手段 1 は、検索対象の文書や検索したい数値表現を入力する手段である。この入力手段 1 は、入力された文書や数値表現を属性付与手段 2 に送る。

属性付与手段 2 は、属性を付与する手段である。この属性付与手段 2 は、属性情報辞書 3 を参照して、入力手段 1 から入力された文書や数値表現の条件部分の前後に属性の“開きタグ”と“閉じタグ”を付与して、数量変換手段 4 に

送る。

属性情報辞書 3 は、属性の情報を格納する辞書である。この属性辞書情報 3 中の属性情報は、属性名と条件の組から構成される。

数量変換手段 4 は、数値情報を変換する手段である。この数量変換手段 4 は、変換情報辞書 5 を参照し、属性付与手段 2 から入力された属性付与後の文書や数値表現に対して、属性の変換前文字列の部分を変換後文字列の部分に変換し、抽出手段 8 に送る。

変換情報辞書 5 は、数量の変換の情報を格納する辞書である。この変換情報辞書 5 中の変換情報は、属性名、変換前文字列及び変換後文字列の組から構成される。

文書格納検索手段 6 は、文書を格納及び検索するための手段である。この文書格納検索手段 6 は、抽出手段 8 から入力された属性付与変換後の文書や、元の文書、検索キーワードを文書データベース 7 に格納し、抽出手段 8 から入力された属性付与後の数値表現が検索キーワードと一致する文書を文書データベース 7 から検索して出力手段 9 に送る。

文書データベース 7 は、検索対象の文書や属性付与変換後の文書を格納するデータベースである。

抽出手段 8 は、キーワードや該当部分を抽出する手段である。この抽出手段 8 は、数量変換手段 4 から入力された変換後の文書や数値表現と、変換後の属性の部分抽出した検索キーワードを、文書格納検索手段 6 に送る。

出力手段 9 は、結果を出力する手段である。この出力手段 9 は、文書格納検索手段 6 から送られた検索結果を出力する。

なお、上述した、属性を付与する処理、文書をデータベース化する処理、文書を格納および検索する処理、並びに該当部分を抽出する処理の各々は、一般的な部分に関しては、既知の自然言語処理技術を用いることで実現できる。

図2は本発明の第1の実施例の数値情報検索装置の動作を説明するフローチャートである。

この図2において、入力手段1で処理を選択し（ステップ201）、投入処理を実行するか（ステップ202）、検索処理を実行するか（ステップ203）、または終了する。

図3は図2のステップ202の投入処理の動作をより具体的に説明したフローチャートであり、図4は図2のステップ203の検索処理の動作をより具体的に説明したフローチャートである。

また、図5は属性情報辞書3の構成を示した図であり、図6は変換情報辞書5の構成を示した図である。なお、図6の変換情報辞書5の変換後文字列における記載「*10**-9m」並びに「*10**-3m」は、それぞれ「 $\times 10^{-9}\text{m}$ 」及び「 $\times 10^{-3}\text{m}$ 」を示している。

図3の投入処理では、まず最初に、検索対象の文書を入力手段1に投入する（ステップ301）。

例えば、以下の例文（a）、（b）また（c）を投入する。

「A社は80ナノメートルの微細加工技術の開発に成功」…（a）

「B社は100nmが限界」…（b）

「0.1mmが限界だった時代は遠く」…（c）

入力手段1に投入された文書は、属性付与手段2に送られる。

次に、属性付与手段2において、属性情報辞書3（詳細構成は、図5参照）を参照して、投入された文書から条件部分を探し、その条件部分に属性を付与する（ステップ302）。

例文（a）では、条件部分として「80ナノメートル」が含まれており、その属性名は「LENGTH」なので、その条件部分の前後に属性の“開きタグ”<LENGTH>および属性の“閉じタグ”</LENGTH>をそれぞれ付与し、

「A社は<LENGTH>80ナノメートル</LENGTH>の微細加工技術の開発に成功」
となる。

例文（b）では、条件部分として「100nm」が含まれており、その属性名は「LENGTH」なので、その条件部分の前後に属性のタグを付与し、

「B社は<LENGTH>100nm</LENGTH>が限界」
となる。

例文（c）では、条件部分として「0.1mm」が含まれているので、同様に、

「<LENGTH>0.1mm</LENGTH>が限界だった時代は遠く」
となる。

属性付与手段2で属性付与された文書は、数量変換手段4に送られる。

次に、数量変換手段4において、変換情報辞書5（詳細構成は、図6参照）を参照して、文書の属性の数量を変換する（ステップ303）。

例文（a）では、属性の数量（：変換前文字列）「ナノメートル」が変換後文字列「 $\times 10^{-9}\text{m}$ 」に変換され、

「A社は<LENGTH>8 0 ×1 0⁻⁹ m</LENGTH>の微細加工技術の開発に成功」
となる。

例文（b）では、変換前文字列「n m」が変換後文字列「×1 0⁻⁹ m」に変換され、

「B社は<LENGTH>1 0 0 ×1 0⁻⁹ m</LENGTH>が限界」
となる。

例文（c）では、変換前文字列「m m」が変換後文字列「×1 0⁻³ m」に変換され、

「<LENGTH>0. 1 ×1 0⁻³ m</LENGTH>が限界だった時代は遠く」
となる。

数量変換手段4で数量変換された文書は、抽出手段8に送られる。

次に、抽出手段8において、数量変換後の属性の部分を検索キーワードとして抽出する（ステップ304）。

例文（a）では、「<LENGTH>8 0 ×1 0⁻⁹ m</LENGTH>」が検索キーワードとして抽出される。

例文（b）では、「<LENGTH>1 0 0 ×1 0⁻⁹ m</LENGTH>」が検索キーワードとして抽出される。

例文（c）では、「<LENGTH>0. 1 ×1 0⁻³ m</LENGTH>」が検索キーワードとして抽出される。

抽出手段8で抽出された検索キーワードは、元の文書及び属性付与変換後の文書とともに文書格納検索手段6に送られる。

最後に、文書格納検索手段 6 によって、文書データベース 7 に、元の文書と属性付与変換後の文書と検索キーワードを格納し（ステップ 3 0 5）、投入処理を終了する。

例文（a）では、元の文書「A社は80ナノメートルの微細加工技術の開発に成功」と、属性付与変換後の文書「A社は<LENGTH>80×10⁻⁹m</LENGTH>の微細加工技術の開発に成功」と、検索キーワード「<LENGTH>80×10⁻⁹m</LENGTH>」とが、文書データベース7に格納される。

例文（b）では、元の文書「B社は100nmが限界」と、属性付与変換後の文書「B社は<LENGTH>100×10⁻⁹m</LENGTH>が限界」と、検索キーワード「<LENGTH>100×10⁻⁹m</LENGTH>」とが、文書データベース7に格納される。

となる。

例文（c）では、元の文書「0.1mmが限界だった時代は遠く」と、属性付与変換後の文書「<LENGTH>0.1×10⁻³m</LENGTH>が限界だった時代は遠く」と、検索キーワード「×10⁻³m」とが、文書データベース7に格納される。

図4の検索処理では、まず最初に、検索したい数値表現を、検索語として入力手段1に入力する（ステップ401）。

例えば、以下の例文（d）を検索語として入力する。

「80nm」…（d）

入力手段1に入力された数値表現（：検索語）は、属性付与手段2に送られる。

次に、属性付与手段 2 において、属性情報辞書 3 を参照して、検索語に属性を付与する（ステップ 4 0 2）。

例文（d）では、その属性名は「LENGTH」なので、その条件部分の前後に属性の“開きタグ” <LENGTH> および属性の“閉じタグ” </LENGTH> をそれぞれ付与し、

「<LENGTH> 8 0 n m</LENGTH>」

となる。

属性付与手段 2 で属性付与された数値表現（：検索語）は、数量変換手段 4 に送られる。

次に、数量変換手段 4 において、変換情報辞書 5 を検索して、検索語の属性の数量を変換する（ステップ 4 0 3）。

例文（d）では、属性の数量（：変換前文字列）「n m」が変換後文字列「 $\times 10^{-9} \text{m}$ 」に変換され、

「<LENGTH> 8 0 $\times 10^{-9} \text{m}$ </LENGTH>」

となる。

数量変換手段 4 で数量変換された検索語は、抽出手段 8 を介して文書格納検索手段 6 に送られる。

次に、文書格納検索手段 6 において、数量変換後の検索語と検索キーワードが一致する文書を、文書データベース 7 に格納されている文書から検索して獲得する（ステップ 4 0 4）。

例文（d）では、検索キーワードが「<LENGTH> 8 0 $\times 10^{-9} \text{m}$ </LENGTH>」である上記例文（a）の文書「A社は80ナノメートルの微細加工技術の開発

に成功」が、数量変換後の検索語「<LENGTH>80×10⁻⁹m</LENGTH>」と検索キーワードが一致する文書として、文書データベース7から検索されて獲得される。

そして、上記ステップ404で獲得した文書を検索結果として出力手段9から出力する（ステップ405）。

なお、出力するのは、属性の部分のみでも、属性の前後の適当な範囲でも、文書全体でもよいし、元の文書からでも、属性付与後の文書からでも、数量変換後の文書からでもよい。

以上のように第1の実施例によれば、文書中に含まれる数値情報を、自動的に統一した表記や単位系へ変換をするので、これら表記や単位系を意識することなく数値情報を検索できる。

第2の実施例

図7は本発明の第2の実施例の、数値情報検索装置のブロック構成図である。なお、図7において、図1と同一あるいは略同一の構成については、同一の符号を付している。

この第2の実施例の数値情報検索装置は、入力手段1、属性付与手段2、属性情報辞書3、数量変換手段4、変換情報辞書5、文書格納検索手段6、文書データベース7、抽出手段8、出力手段9、並べ替え手段10、及び抽出情報辞書10を備える。即ち、この第2の実施例は、上記図1に示した第1の実施例の数値情報検索装置に、並べ替え手段10及び抽出情報辞書11を追加したものである。

並べ替え手段 10 は、抽出結果を並べ替えるための手段である。抽出情報辞書 11 は、抽出と並べ替えの情報を格納する辞書である。抽出情報辞書 11 中の抽出情報は、属性名、範囲指定語、比較条件及び整列順の組から構成される。

並べ替え手段 10 は、抽出情報辞書 11 を参照して、文書格納検索手段 6 から送られた検索キーワードを抽出し、整列順に従って並べ替えて、文書データベース 7 から検索された文書を出力手段 9 に送る。

この第 2 の実施例の数値情報検索装置の全体の動作及び投入処理の動作は、上記第 1 の実施例の全体動作（図 2 参照）および投入処理の動作（図 3 参照）と同様である。よって、第 2 の実施例の検索処理の動作についてのみ、以下に説明する。

図 8 は本発明の第 2 の実施例の数値情報検索装置における検索処理の動作を説明するフローチャートである。また、図 9 は抽出情報辞書 11 の構成を示した図である。なお、図 9 の抽出情報辞書 11 の比較条件の記述である「<=」及び「==」は、それぞれ「 \leq 」及び「=」を表す。

なお、以下の第 2 の実施例の検索処理の説明においては、上記第 1 の実施例の動作の説明に用いた例文（a）、（b）および（c）の投入処理がすでになされているものとする。

図 8 に示された検索処理では、まず最初に、検索したい数値表現を、検索語として入力手段 1 に入力する（ステップ 801）。

例えば、以下の例文（e）を検索語として入力する。

「100nm以下」…（e）

入力手段 1 に入力された数値表現（：検索語）は、属性付与手段 2 に送られ

る。

次に、属性付与手段 2 において、属性情報辞書 3 を参照して、検索語から条件部分を探し、その条件部分に属性を付与する（ステップ 8 0 2）。

例文（e）では、条件部分として「1 0 0 n m」が含まれており、その属性名は「LENGTH」なので、その条件部分の前後に属性の“開きタグ” <LENGTH> および属性の“閉じタグ” </LENGTH> をそれぞれ付与し、

「<LENGTH> 1 0 0 n m</LENGTH>以下」

となる。

属性付与手段 2 で属性付与された数値表現（：検索語）は、数量変換手段 4 に送られる。

次に、数量変換手段 4 において、変換情報辞書 5 を参照して、検索語の属性の数量を変換する（ステップ 4 0 3）。

例文（e）では、属性の数量（：変換前文字列）「n m」が変換後文字列「 $\times 10^{-9}\text{m}$ 」に変換され、

「<LENGTH> 1 0 0 $\times 10^{-9}\text{m}$ </LENGTH>以下」

となる。

数量変換手段 4 で数量変換された検索語は、抽出手段 8 に送られる。

次に、抽出手段 8 において、検索語の数量変換後の属性名と数値と範囲指定語を抽出する（ステップ 8 0 4）。

例文（e）では、属性名として「LENGTH」が抽出され、数値として「1 0 0 $\times 10^{-9}\text{m}$ 」が抽出され、範囲指定語として「以下」が抽出される。

抽出手段 8 で抽出された属性名と数値と範囲指定語は、文書格納検索手段 6

に送られる。

次に、文書格納検索手段6において、検索語から抽出された属性名と属性名が一致する検索キーワードを、文書データベース7に格納されている検索キーワードから検索する（ステップ805）。

例文（e）では、属性名が「LENGTH」である検索キーワードの検索がなされ、上記例文（a）の検索キーワード「<LENGTH>80×10⁻⁹m</LENGTH>」、上記例文（b）の検索キーワード「<LENGTH>100×10⁻⁹m</LENGTH>」、及び上記例文（c）の検索キーワード「<LENGTH>0.1×10⁻³m</LENGTH>」が、文書データベース7から検索される。

文書データベース7から検索された検索キーワードは、属性付与変換後の検索語とともに並べ替え手段10に送られる。

次に、並べ替え手段10において、抽出情報辞書11（詳細構成は図9参照）を参照して、検索語の数値と文書データベース7から検索された検索キーワードの数値とを比較する（ステップ806）。

例文（e）では、検索語の数値は「100×10⁻⁹m」であって、

i) 上記例文（a）の検索キーワードの数値「80×10⁻⁹m」との比較では、

$$80 \times 10^{-9} \text{ m} \text{ 「<」 } 100 \times 10^{-9} \text{ m}$$

であり、

ii) 上記例文（b）の検索キーワードの数値「100×10⁻⁹m」との比較では、

$$100 \times 10^{-9} \text{ m} \text{ 「=」 } 100 \times 10^{-9} \text{ m}$$

であり、

iii) 上記例文 (b) の検索キーワードの数値「 $100 \times 10^{-9} \text{m}$ 」との比較では、

$$0.1 \times 10^{-3} \text{m} \text{「<」} 100 \times 10^{-9} \text{m}$$

である。

次に、並べ替え手段 10 において、抽出情報辞書 11 を参照して、比較した結果が範囲指定語の比較条件に適合する検索キーワードを選択する (ステップ 807)。

例文 (e) では、範囲指定語「以下」という比較条件に適合するのは、比較結果が「<」である上記例文 (a) の検索キーワード「<LENGTH> $80 \times 10^{-9} \text{m}$ </LENGTH>」と、比較結果が「=」である上記例文 (b) の検索キーワード「<LENGTH> $100 \times 10^{-9} \text{m}$ </LENGTH>」であり、これら 2 つの検索キーワードが選択される。

次に、並べ替え手段 10 において、抽出情報辞書 11 を参照して、選択した検索キーワードを整列順に従って並べ替える (ステップ 808)。

例文 (e) では、その範囲指定語「以下」の整列順は「昇順」に設定されているので (図 9 参照)、上記 2 つの検索キーワードは、

「<LENGTH> $80 \times 10^{-9} \text{m}$ </LENGTH>」 (: 上記例文 (a) の検索キーワード)

「<LENGTH> $100 \times 10^{-9} \text{m}$ </LENGTH>」 (: 上記例文 (b) の検索キーワード)

の順で並べ替えられる。

並べ替えられた検索キーワードは、文書格納検索手段6に送られる。

次に、文書格納検索手段6において、並べ替えられた検索キーワードと検索キーワードが一致する文書を、文書データベース7に格納している文書から検索して獲得する（ステップ809）。

そして、上記ステップ809で獲得した文書を、並べ替えられた検索キーワードの順に従って、検索結果として出力手段9から出力する（ステップ810）。

上記例文（e）では、

「A社は80ナノメートルの微細加工技術の開発に成功」（：上記例文（a））

「B社は100nmが限界」（：上記例文（b））

の順で出力される。

なお、以下で例文（f）の検索語として入力した場合には、範囲指定語は「かそれ以下」となる（図9参照）。

「100nmかそれ以下」…（f）

この例文（f）では、範囲指定語「かそれ以下」の整列順は「降順」に設定されているので（図9参照）、

「<LENGTH>100×10⁻⁹m<LENGTH>」（：上記例文（b）の検索キーワード）

「<LENGTH>80×10⁻⁹m<LENGTH>」（：上記例文（a）の検索キーワード）

の順に検索キーワードが並べ替えられて、

「B社は100nmが限界」（：上記例文（b））

「A社は80ナノメートルの微細加工技術の開発に成功」（：上記例文（a））

の順で出力され、範囲指定語が「以下」の場合とは逆となる。

以上のように第2の実施例によれば、範囲の表現を数値の大小関係に変換して比較することによって、数式を意識しなくても数値情報を検索することができる。また、範囲の表現に従って検索結果を並べ替えることにより、効率よく検索結果を確認できる。更に、通常 of 自然言語 of 検索に適用することによって、数値表現 of 検索が容易になる。

なお、上記第1の実施例及び第2の実施例では、属性を付与する方法として、属性タグを原文に挿入する方法を用いたが、本発明では別の方法を用いてもよい。例えば、原文を変更することなく、「原文の35文字目から40文字目までが「LENGTH」の属性を持っている」といった情報が記載された“原文の属性情報データ”として、原文とは独立して保持する方法も考えられる。

また上記第1の実施例及び第2の実施例では、数値情報のみを検索語および検索対象とする数値情報検索装置を説明したが、本発明は、数値以外の文言を検索語または検索対象とする検索方法や検索装置と組み合わせて使用することも可能である。

以上のように本発明によれば、数値的な変換をすることによって、表記や単位系を意識しなくても数値情報を検索できるという効果がある。

What is claimed is:

1. 自然言語で数値情報を検索する数値情報検索装置において、
検索対象の文書や検索したい数量表現を入力する入力手段と、
属性を表す属性名と属性を付与するための条件とが格納された属性情報辞書と、
上記入力された文書や数値表現に上記属性情報辞書を参照して属性を付与する属性付与手段と、
属性を表す属性名と変換前文字列と変換後文字列とが格納された変換情報辞書と、
上記属性付与後の文書や数値表現に対して、上記変換辞書情報を参照して、属性付与部分の数量を変換前文字列から変換後文字列に変換することによって、前記属性付与部分の数値情報を変換する数量変換手段と
を備えたことを特徴とする数値情報検索装置。
2. 請求項 1 記載の数値情報検索装置において、
上記属性付与変換後の文書から特定の属性部分を検索キーワードとして抽出する抽出手段と、
文書データが格納される文書データベースと、
上記属性付与変換後の文書と上記入力された元の文書と上記抽出された検索キーワードとを上記文書データベースに格納する文書格納検索手段と
を更に備えたことを特徴とする数値情報検索装置。
3. 請求項 2 記載の数値情報検索装置において、
検索結果を出力する出力手段を更に備え、

上記文書格納検索手段は、上記属性付与変換後の数値表現と検索キーワードが一致する文書を上記文書データベースから検索し、上記検索結果として上記出力手段に送ることを特徴とする数値情報検索装置。

4. 請求項2記載の数値情報検索装置において、

属性を表す属性名と範囲の表現を表す範囲指定語と比較の大小関係を表す比較条件の出力の並べ替えのための順序を表す整列順とが格納された抽出情報辞書と、前記出力の並べ替えのための並べ替え手段と、検索結果を出力する出力手段とを更に備え、

上記抽出手段は、上記属性付与変換後の数値表現から属性名と数値と範囲指定語とを抽出し、

上記並べ替え手段は、上記数値表現から抽出された数値を、上記数値表現から抽出された属性名と属性名が一致する複数の検索キーワードの数値とそれぞれ比較し、上記抽出情報辞書を参照して、上記比較結果が範囲指定語の比較条件に適合する検索キーワードを選択し、選択した検索キーワードを整列順に従って並べ替え、

上記文書格納検索手段は、上記数値表現から抽出された属性名と属性名が一致する検索キーワードを上記文書データベースから検索して上記並べ替え手段に送り、上記並べ替えされた検索キーワードと一致する文書を上記文書データベースから検索し、これらの検索した文書を上記並べ替えされた検索キーワードの順序に従って上記検索結果として上記出力手段に送ることを特徴とする数値情報検索装置。

5. 請求項1記載の数値情報検索装置において、

上記属性付与手段は、上記属性情報辞書を参照して、属性の開きタグと閉じタグを原文の条件部分の前後に付与することを特徴とする数値情報検索装置。

6. 請求項 1 記載の数値情報検索装置において、

上記属性付与手段は、上記属性情報辞書を参照して、原文を変更することなく、原文中の条件部分の属性を示す属性情報データを、上記原文とは別に保持できるように付与することを特徴とする数値情報検索装置。

ABSTRACT

表記や単位系を意識しなくても数値情報を検索することができる数値情報検索装置を実現するために、本発明の数値情報検索装置は、検索対象の文書や検索したい数量表現を入力する入力手段 1, 属性を表す属性名と属性を付与するための条件とが格納された属性情報辞書 3, 上記入力された文書や数値表現に上記属性情報辞書 3 を参照して属性を付与する属性付与手段 2, 属性を表す属性名と変換前文字列と変換後文字列とが格納された変換情報辞書 5, 及び上記属性付与後の文書や数値表現に対して、上記変換情報辞書 5 を参照して、属性付与部分の数量を変換前文字列から変換後文字列に変換することによって、上記属性付与部分の数値情報を変換する数量変換手段 4 とを備える。